

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10/565,954
Source: 1FW16
Date Processed by STIC: 10/3/06

ENTERED



IFW16

RAW SEQUENCE LISTING

DATE: 10/03/2006

PATENT APPLICATION: US/10/565,954

TIME: 10:33:01

Input Set : A:\GC812-PCT-SEQLIST.TXT

Output Set: N:\CRF4\10032006\J565954.raw

4 <110> APPLICANT: Genencor International, Inc.
 5 Fenel, Fred
 6 Clarkson, Kathleen A.
 8 <120> TITLE OF INVENTION: Modified Enzymes, Methods to Produce
 9 Modified Enzymes and Uses Thereof
 12 <130> FILE REFERENCE: GC812-PCT
 C--> 14 <140> CURRENT APPLICATION NUMBER: US/10/565,954
 C--> 15 <141> CURRENT FILING DATE: 2006-07-31
 17 <150> PRIOR APPLICATION NUMBER: US 60/503,251
 18 <151> PRIOR FILING DATE: 2003-09-15
 20 <160> NUMBER OF SEQ ID NOS: 51
 22 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 24 <210> SEQ ID NO: 1
 25 <211> LENGTH: 223
 26 <212> TYPE: PRT
 27 <213> ORGANISM: Trichoderma reesei
 29 <400> SEQUENCE: 1
 30 Met Val Ser Phe Thr Ser Leu Leu Ala Gly Val Ala Ala Ile Ser Gly
 31 1 5 10 15
 32 Val Leu Ala Ala Pro Ala Ala Glu Val Glu Ser Val Ala Val Glu Lys
 33 20 25 30
 34 Arg Gln Thr Ile Gln Pro Gly Thr Gly Tyr Asn Asn Gly Tyr Phe Tyr
 35 35 40 45
 36 Ser Tyr Trp Asn Asp Gly His Gly Gly Val Thr Tyr Thr Asn Gly Pro
 37 50 55 60
 38 Gly Gly Gln Phe Ser Val Asn Trp Ser Asn Ser Gly Asn Phe Val Gly
 39 65 70 75 80
 40 Gly Lys Gly Trp Gln Pro Gly Thr Lys Asn Lys Val Ile Asn Phe Ser
 41 85 90 95
 42 Gly Ser Tyr Asn Pro Asn Gly Asn Ser Tyr Leu Ser Val Tyr Gly Trp
 43 100 105 110
 44 Ser Arg Asn Pro Leu Ile Glu Tyr Tyr Ile Val Glu Asn Phe Gly Thr
 45 115 120 125
 46 Tyr Asn Pro Ser Thr Gly Ala Thr Lys Leu Gly Glu Val Thr Ser Asp
 47 130 135 140
 48 Gly Ser Val Tyr Asp Ile Tyr Arg Thr Gln Arg Val Asn Gln Pro Ser
 49 145 150 155 160
 50 Ile Ile Gly Thr Ala Thr Phe Tyr Gln Tyr Trp Ser Val Arg Arg Asn
 51 165 170 175
 52 His Arg Ser Ser Gly Ser Val Asn Thr Ala Asn His Phe Asn Ala Trp
 53 180 185 190
 54 Ala Gln Gln Gly Leu Thr Leu Gly Thr Met Asp Tyr Gln Ile Val Ala
 55 195 200 205

see p. 6

RAW SEQUENCE LISTING

DATE: 10/03/2006

PATENT APPLICATION: US/10/565,954

TIME: 10:33:01

Input Set : A:\GC812-PCT-SEQLIST.TXT

Output Set: N:\CRF4\10032006\J565954.raw

```

56 Val Glu Gly Tyr Phe Ser Ser Gly Ser Ala Ser Ile Thr Val Ser
57      210                      215                      220
59 <210> SEQ ID NO: 2
60 <211> LENGTH: 781
61 <212> TYPE: DNA
62 <213> ORGANISM: Trichoderma reesei
64 <400> SEQUENCE: 2
65 atggtctcct tcacctccct cctcgccggc gtcgccgcca tctcgggcgt cttggccgct      60
66 cccgccgccc aggtcgaatc cgtggctgtg gagaagcgcc agacgattca gcccggcacg      120
67 ggctacaaca acggctactt ctactcgtac tggaacgatg gccacggcgg cgtgacgtac      180
68 accaatgggtc ccggcgggca gttctccgtc aactgggtcca actcgggcaa ctttgcggc      240
69 ggcaagggat ggcagcccgg caccaagaac aagtaagact acctactctt accccctttg      300
70 accaaccacag cacaacacaa tacaacacat gtgactacca atcatggaat cggatctaac      360
71 agctgtgttt tcaaaaaaaaa gggcatcaa cttctcgggc agctacaacc ccaacggcaa      420
72 cagctacctc tccgtgtacg gctggccccg caacccccctg atcgagtact acatcgtcga      480
73 gaactttggc acctacaacc cgtccacggg cgccaccaag ctgggcgagg tcacctccga      540
74 cggcagcgtc tacgacattt accgcacgca gcgcgtcaac cagccgtcca tcatcggcac      600
75 cgccaccttt taccagtact ggtccgtccg ccgcaaccac cgctcgagcg gctccgtcaa      660
76 cggggcgaaac caattcaacg cgtggggtca gcaaggcctg acgctcggga cgatggatta      720
77 ccagattggt gccgtggagg gttactttag ctctggctct gttccatca ccgtcagcta      780
78 a                                                                    781
80 <210> SEQ ID NO: 3
81 <211> LENGTH: 234
82 <212> TYPE: PRT
83 <213> ORGANISM: Trichoderma reesei
85 <400> SEQUENCE: 3
86 Met Lys Phe Leu Gln Val Leu Pro Ala Leu Ile Pro Ala Ala Leu Ala
87 1          5          10          15
88 Gln Thr Ser Cys Asp Gln Trp Ala Thr Phe Thr Gly Asn Gly Tyr Thr
89      20          25          30
90 Val Ser Asn Asn Leu Trp Gly Ala Ser Ala Gly Ser Gly Phe Gly Cys
91      35          40          45
92 Val Thr Ala Val Ser Leu Ser Gly Gly Ala Ser Trp His Ala Asp Trp
93      50          55          60
94 Gln Trp Ser Gly Gly Gln Asn Asn Val Lys Ser Tyr Gln Asn Ser Gln
95 65          70          75          80
96 Ile Ala Ile Pro Gln Lys Arg Thr Val Asn Ser Ile Ser Ser Met Pro
97      85          90          95
98 Thr Thr Ala Ser Trp Ser Tyr Ser Gly Ser Asn Ile Arg Ala Asn Val
99      100         105         110
100 Ala Tyr Asp Leu Phe Thr Ala Ala Asn Pro Asn His Val Thr Tyr Ser
101      115         120         125
102 Gly Asp Tyr Glu Leu Met Ile Trp Leu Gly Lys Tyr Gly Asp Ile Gly
103      130         135         140
104 Pro Ile Gly Ser Ser Gln Gly Thr Val Asn Val Gly Gly Gln Ser Trp
105 145         150         155         160
106 Thr Leu Tyr Tyr Gly Tyr Asn Gly Ala Met Gln Val Tyr Ser Phe Val
107      165         170         175
108 Ala Gln Thr Asn Thr Thr Asn Tyr Ser Gly Asp Val Lys Asn Phe Phe

```

RAW SEQUENCE LISTING

DATE: 10/03/2006

PATENT APPLICATION: US/10/565,954

TIME: 10:33:01

Input Set : A:\GC812-PCT-SEQLIST.TXT

Output Set: N:\CRF4\10032006\J565954.raw

```

109          180          185          190
110 Asn Tyr Leu Arg Asp Asn Lys Gly Tyr Asn Ala Ala Gly Gln Tyr Val
111          195          200          205
112 Leu Ser Tyr Gln Phe Gly Thr Glu Pro Phe Thr Gly Ser Gly Thr Leu
113          210          215          220
114 Asn Val Ala Ser Trp Thr Ala Ser Ile Asn
115 225          230
117 <210> SEQ ID NO: 4
118 <211> LENGTH: 826
119 <212> TYPE: DNA
120 <213> ORGANISM: Trichoderma reesei
122 <400> SEQUENCE: 4
123 atgaagtcc ttcaagtcc cctgccctc ataccggccg ccctggccca aaccagctgt      60
124 gaccagtggg caaccttcac tggcaacggc tacacagtca gcaacaacct ttggggagca      120
125 tcagccggct ctggatttgg ctgctgacg gcggtatcgc tcagcggcgg ggccctctgg      180
126 cagcgagact ggcagtggtc cggcggccag aacaacgtca agtcgtacca gaactctcag      240
127 attgccattc cccagaagag gaccgtcaac agcatcagca gcatgccac cactgccagc      300
128 tggagctaca gcgggagcaa catccgcgct aatgttgcgt atgacttggt caccgcagcc      360
129 aaccggaatc atgtcacgta ctcgggagac tacgaactca tgatctggta agccataaga      420
130 agtgaccctc cttgatagtt tcgactaaca acatgtcttg aggcttggca aatacggcga      480
131 tattgggccg attgggtcct cacagggaac agtcaacgtc ggtggccaga gctggacgct      540
132 ctactatggc tacaacggag ccatgcaagt ctattccttt gtggcccaga ccaacactac      600
133 caactacagc ggagatgtca agaacttctt caattatctc cgagacaata aaggatacaa      660
134 cgctgcaggc caatatgttc ttagtaagtc accctcactg tgactgggct gagtttggtg      720
135 caacgtttgc taacaaaacc ttcgtatagg ctaccaattt ggtaccgagc ccttcacggg      780
136 cagtggaaact ctgaacgtcg catcctggac cgcactatc aactaa      826
138 <210> SEQ ID NO: 5
139 <211> LENGTH: 222
140 <212> TYPE: PRT
141 <213> ORGANISM: Trichoderma reesei
143 <400> SEQUENCE: 5
144 Met Val Ser Phe Thr Ser Leu Leu Ala Ala Ser Pro Pro Ser Arg Ala
145 1          5          10          15
146 Ser Cys Arg Pro Ala Ala Glu Val Glu Ser Val Ala Val Glu Lys Arg
147          20          25          30
148 Gln Thr Ile Gln Pro Gly Thr Gly Tyr Asn Asn Gly Tyr Phe Tyr Ser
149          35          40          45
150 Tyr Trp Asn Asp Gly His Gly Gly Val Thr Tyr Thr Asn Gly Pro Gly
151          50          55          60
152 Gly Gln Phe Ser Val Asn Trp Ser Asn Ser Gly Asn Phe Val Gly Gly
153 65          70          75          80
154 Lys Gly Trp Gln Pro Gly Thr Lys Asn Lys Val Ile Asn Phe Ser Gly
155          85          90          95
156 Ser Tyr Asn Pro Asn Gly Asn Ser Tyr Leu Ser Val Tyr Gly Trp Ser
157          100          105          110
158 Arg Asn Pro Leu Ile Glu Tyr Tyr Ile Val Glu Asn Phe Gly Thr Tyr
159          115          120          125
160 Asn Pro Ser Thr Gly Ala Thr Lys Leu Gly Glu Val Thr Ser Asp Gly
161          130          135          140

```

RAW SEQUENCE LISTING

DATE: 10/03/2006

PATENT APPLICATION: US/10/565,954

TIME: 10:33:01

Input Set : A:\GC812-PCT-SEQLIST.TXT

Output Set: N:\CRF4\10032006\J565954.raw

```

162 Ser Val Tyr Asp Ile Tyr Arg Thr Gln Arg Val Asn Gln Pro Ser Ile
163 145 150 155 160
164 Ile Gly Thr Ala Thr Phe Tyr Gln Tyr Trp Ser Val Arg Arg Asn His
165 165 170 175
166 Arg Ser Ser Gly Ser Val Asn Thr Ala Asn His Phe Asn Ala Trp Ala
167 180 185 190
168 Gln Gln Gly Leu Thr Leu Gly Thr Met Asp Tyr Gln Ile Val Ala Val
169 195 200 205
170 Glu Gly Tyr Phe Ser Ser Gly Ser Ala Ser Ile Thr Val Ser
171 210 215 220
173 <210> SEQ ID NO: 6
174 <211> LENGTH: 227
175 <212> TYPE: PRT
176 <213> ORGANISM: Humicola insolens
178 <400> SEQUENCE: 6
179 Met Val Ser Leu Lys Ser Val Leu Ala Ala Ala Thr Ala Val Ser Ser
180 1 5 10 15
181 Ala Ile Ala Ala Pro Phe Asp Phe Val Pro Arg Asp Asn Ser Thr Ala
182 20 25 30
183 Leu Gln Ala Arg Gln Val Thr Pro Asn Ala Glu Gly Trp His Asn Gly
184 35 40 45
185 Tyr Phe Tyr Ser Trp Trp Ser Asp Gly Gly Gly Gln Val Gln Tyr Thr
186 50 55 60
187 Asn Leu Glu Gly Ser Arg Tyr Gln Val Arg Trp Arg Asn Thr Gly Asn
188 65 70 75 80
189 Phe Val Gly Gly Lys Gly Trp Asn Pro Gly Thr Gly Arg Thr Ile Asn
190 85 90 95
191 Tyr Gly Gly Tyr Phe Asn Pro Gln Gly Asn Gly Tyr Leu Ala Val Tyr
192 100 105 110
193 Gly Trp Thr Arg Asn Pro Leu Val Glu Tyr Tyr Val Ile Glu Ser Tyr
194 115 120 125
195 Gly Thr Tyr Asn Pro Gly Ser Gln Ala Gln Tyr Lys Gly Thr Phe Tyr
196 130 135 140
197 Thr Asp Gly Asp Gln Tyr Asp Ile Phe Val Ser Thr Arg Tyr Asn Gln
198 145 150 155 160
199 Pro Ser Ile Asp Gly Thr Arg Thr Phe Gln Gln Tyr Trp Ser Ile Arg
200 165 170 175
201 Lys Asn Lys Arg Val Gly Gly Ser Val Asn Met Gln Asn His Phe Asn
202 180 185 190
203 Ala Trp Gln Gln His Gly Met Pro Leu Gly Gln His Tyr Tyr Gln Val
204 195 200 205
205 Val Ala Thr Glu Gly Tyr Gln Ser Ser Gly Glu Ser Asp Ile Tyr Val
206 210 215 220
207 Gln Thr His
208 225
210 <210> SEQ ID NO: 7
211 <211> LENGTH: 210
212 <212> TYPE: PRT
213 <213> ORGANISM: Bacillus stearothermophilus

```

RAW SEQUENCE LISTING

DATE: 10/03/2006

PATENT APPLICATION: US/10/565,954

TIME: 10:33:01

Input Set : A:\GC812-PCT-SEQLIST.TXT

Output Set: N:\CRF4\10032006\J565954.raw

215 <400> SEQUENCE: 7

```

216 Met Lys Leu Lys Lys Lys Met Leu Thr Leu Leu Leu Thr Ala Ser Met
217 1 5 10 15
218 Ser Phe Gly Leu Phe Gly Ala Thr Ser Ser Ala Ala Thr Asp Tyr Trp
219 20 25 30
220 Gln Tyr Trp Thr Asp Gly Gly Gly Met Val Asn Ala Val Asn Gly Pro
221 35 40 45
222 Gly Gly Asn Tyr Ser Val Thr Trp Gln Asn Thr Gly Asn Phe Val Val
223 50 55 60
224 Gly Lys Gly Trp Thr Val Gly Ser Pro Asn Arg Val Ile Asn Tyr Asn
225 65 70 75 80
226 Ala Gly Ile Trp Glu Pro Ser Gly Asn Gly Tyr Leu Thr Leu Tyr Gly
227 85 90 95
228 Trp Thr Arg Asn Ala Leu Ile Glu Tyr Tyr Val Val Asp Ser Trp Gly
229 100 105 110
230 Thr Tyr Arg Pro Thr Gly Asn Tyr Lys Gly Thr Val Asn Ser Asp Gly
231 115 120 125
232 Gly Thr Tyr Asp Ile Tyr Thr Thr Met Arg Tyr Asn Ala Pro Ser Ile
233 130 135 140
234 Asp Gly Thr Gln Thr Phe Gln Gln Phe Trp Ser Val Arg Gln Ser Lys
235 145 150 155 160
236 Arg Pro Thr Gly Ser Asn Val Ser Ile Thr Phe Ser Asn His Val Asn
237 165 170 175
238 Ala Trp Arg Ser Lys Gly Met Asn Leu Gly Ser Ser Trp Ala Tyr Gln
239 180 185 190
240 Val Leu Ala Thr Glu Gly Tyr Gln Ser Ser Gly Arg Ser Asn Val Thr
241 195 200 205
242 Val Trp
243 210

```

245 <210> SEQ ID NO: 8

246 <211> LENGTH: 229

247 <212> TYPE: PRT

248 <213> ORGANISM: Trichoderma reesei

250 <400> SEQUENCE: 8

```

251 Met Val Ala Phe Ser Ser Leu Ile Cys Ala Leu Thr Ser Ile Ala Ser
252 1 5 10 15
253 Thr Leu Ala Met Pro Thr Gly Leu Glu Pro Glu Ser Ser Val Asn Val
254 20 25 30
255 Thr Glu Arg Gly Met Tyr Asp Phe Val Leu Gly Ala His Asn Asp His
256 35 40 45
257 Arg Arg Arg Ala Ser Ile Asn Tyr Asp Gln Asn Tyr Gln Thr Gly Gly
258 50 55 60
259 Gln Val Ser Tyr Ser Pro Ser Asn Thr Gly Phe Ser Val Asn Trp Asn
260 65 70 75 80
261 Thr Gln Asp Asp Phe Val Val Gly Val Gly Trp Thr Thr Gly Ser Ser
262 85 90 95
263 Ala Pro Ile Asn Phe Gly Gly Ser Phe Ser Val Asn Ser Gly Thr Gly
264 100 105 110
265 Leu Leu Ser Val Tyr Gly Trp Ser Thr Asn Pro Leu Val Glu Tyr Tyr

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/565,954

DATE: 10/03/2006
TIME: 10:33:02

Input Set : A:\GC812-PCT-SEQLIST.TXT
Output Set: N:\CRF4\10032006\J565954.raw

PJT
Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:28; Xaa Pos. 2,5,6,7,8,9,10,13,14,15,18,21,22,23,24,25,28,32,33,34,35
Seq#:28; Xaa Pos. 36,37,38,39,40,41,42,43,45,46,47,49,54,55,56,61,62,63,64
Seq#:28; Xaa Pos. 65,66,67,68,69,70,71,72,73,74,76,78,80,81,82,83,84,85,87
Seq#:28; Xaa Pos. 89,91,92,93,94,104,105,106,108,110,111,112,113,115,118
Seq#:28; Xaa Pos. 123,124,125,128,129,134,135,136,137,139,140,141,142,143
Seq#:28; Xaa Pos. 144,146,147,149,151,152,154,157,159,161,166,174,177,178
Seq#:28; Xaa Pos. 179,181,182,185,186,187,188,190,192,193,194,198,199,200
Seq#:28; Xaa Pos. 202,210,211,213,214,215,216,217,218,219,220

VERIFICATION SUMMARY

DATE: 10/03/2006

PATENT APPLICATION: US/10/565,954

TIME: 10:33:02

Input Set : A:\GC812-PCT-SEQLIST.TXT

Output Set: N:\CRF4\10032006\J565954.raw

L:14 M:270 C: Current Application Number differs, Replaced Current Application Number
L:15 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:1036 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:28
L:1041 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:1045 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:28
L:1046 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:0
L:1048 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:16
L:1050 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:32
L:1052 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:48
L:1054 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:64
L:1056 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:80
L:1058 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:96
L:1060 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:112
L:1062 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:128
L:1064 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:144
L:1066 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:160
L:1068 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:176
L:1070 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:192
L:1072 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:208